B1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00819 A1

(51) International Patent Classification[7]: C12N 15/10, 15/90, 15/67

(21) International Application Number: PCT/US00/17154

(22) International Filing Date: 21 June 2000 (21.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/141,103 25 June 1999 (25.06.1999) US

(71) Applicants *(for all designated States except US)*: **E.I. DU PONT DE NEMOURS AND COMPANY** [US/US]; 1007 Market Street, Wilmington, DE 19898 (US). **GENENCOR INTERNATIONAL, INC.** [US/US]; 200 Meridian Centre Boulevard, Rochester, NY 14618 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **CHAUHAN, Sarita** [IN/US]; 384 Egypt Run Road, Landenberg, PA 19350 (US).

(74) Agent: **FLOYD, Linda, Axamethy**; E.I. Du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States *(national)*: CA, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/00819 A1

(54) Title: TARGETED GENE REPLACEMENTS IN ENTERIC BACTERIA USING LINEAR DNA

(57) Abstract: The present invention provides a rapid and specific method for targeted gene replacement in bacteria. The modified or disrupted genes or other nucleic acid fragments are transformed into their desired hosts and the transformants are screened for desired integrations using a selectable marker for an altered phenotype. More specifically, the instant invention creates site-directed mutants of enteric bacteria by allelic exchange to identify an altered phenotype that will define the function or effect of the mutated gene. This invention also describes how linear DNA fragments harboring a functional gene can be used for replacement of a mutant gene to restore gene functionality.

TITLE

TARGETED GENE REPLACEMENTS IN ENTERIC BACTERIA USING LINEAR DNA

FIELD OF THE INVENTION

The present invention relates to the field of molecular biology and microbiology. More specifically, methods have been developed for the introduction of site-directed alterations in the *Escherichia coli* genome using linear DNA molecules to generate stable transformants.

BACKGROUND OF THE INVENTION

With the advent of large scale genome sequencing efforts, enormous amounts of sequence information are being made available to the research community. Genome sequencing efforts have been completed for several prokaryotes including *Escherichia coli*. At least 46 genomes of different genera of bacteria of industrial, pharmaceutical and agricultural importance, have been completely sequenced and about 70 are in progress (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/bact.html). However, only a small portion of the genes sequenced from all these efforts have been functionally characterized.

The need for discovery and analyses of gene function has spawned several new areas of research referred to as functional genomics, proteomics, and metabolomics. Functional genomics seeks to discover gene function once nucleotide sequence information is available. Proteomics (the study of protein properties such as expression, post-translational modifications, interactions, etc.) and metabolomics (analysis of metabolite pools) are fast-emerging fields complementing functional genomics, that provide a global, integrated view of cellular processes. The variety of techniques and methods used in this effort include the use of bioinformatics, gene-array chips, mRNA differential display, disease models, protein discovery and expression, and target validation. The ultimate goal of many of these efforts has been to develop high-throughput screens for genes of unknown function.

A variety of techniques have been applied to elucidate gene function including identifying the interacting protein partner of a gene product as in the yeast 2-hybrid system (Bolger, G., *Methods Mol. Biol.* 88:101-131 (1998)) and transposon tagging which is useful in both microbial and eukaryotic genomes (Kumar et al., *Plant Biotechnol.* 15:159-165 (1998). The logarithmic increase in sequence data has driven the need for high-throughput (HTP) functional genomics screens. However, relatively few HTP methods have been developed to date.

Traditional methods for the determination of gene function still remain the basis of the functional genomics effort.

Classically, the first and most basic analysis for any gene is to assess the phenotype of the organism when the gene of interest is altered or rendered nonfunctional. The impact of such alterations as well as knockouts on proteome or metabolome can be measured to understand gene function. The "null mutations" are often constructed by gene disruption (also called gene knockouts) by homologous recombination resulting in allelic exchange or gene replacement. Gene disruption vectors are constructed by recombinant DNA techniques. Upon transformation into the organism the DNA construct with disrupted gene integrates at the resident location in the genome by homologous recombination and replaces the functional copy of the gene with the nonfunctional disrupted gene. Gene disruption vectors are constructed from a genomic clone containing the gene of interest, using a vector that does not replicate in the host organism to be mutated. The final disruption construct has 1) a selectable marker inserted in and/or around the gene of interest in a manner that eliminates expression of the gene of interest or renders non-functional any protein product that is produced, and 2) additional unmodified genomic DNA flanking the gene disruption site of interest that helps target integration events to the disruption site. Gene disruption vectors have been constructed for microbial and mammalian systems by (i) identifying and isolating a genomic clone containing the gene of interest; (ii) determining convenient restriction sites within or flanking the gene; (iii) subcloning a selectable marker into the gene at the restriction sites thereby inactivating the gene; (iv) transforming the altered gene containing the marker into the host organism; and (v) screening the transformant for an altered phenotype (see for example Sedivy et al., *Proc. Natl. Acad. Sci. USA* 86:227-31 (1989)).

The above process has worked well for disruption of genes in a range of organisms but has several inherent limitations. First, a genomic clone of suitable length must be isolated from a suitable library, most often by hybridization or cloned by PCR into a suitable plasmid. Second, the methodology is completely limited to existing restriction sites which must be compatible with the selectable marker to be ligated and must occur in positions such that ligation of the selectable marker makes the gene of interest nonfunctional. Additional restriction sites that occur outside the gene of interest can make cloning of the selectable marker into the genomic DNA very difficult. Third, because this process relies on traditional restriction analysis and cloning, it is inherently a labor intensive, one-gene-at-a-time approach. Fourth, cosuppression or quelling (Cogoni et al., *EMBO J.* 15:3153-3163 (1996)), a gene silencing phenomenon seen in fungi and plants

can result when a gene disruption vector leaves the promoter and part of the 5' coding region intact. Fifth, these methods often give unpredictable results. For example, a desired gene targeting will result in replacement of the functional allele with a null mutation results. If, however, non-targeted, ectopic integration occurs (and this undesired integration often predominates) then the transformant harbors both a functional and nonfunctional copy of the gene. Yeast and *Escherichia coli* have previously been used for *in vivo* cloning for double strand break repair or gap repair (Oldenber et al., *Nucleic Acids Research.* 25:451-452 (1997); Hua et al., *Plasmid* 38:91-96 (1997); Oliner et al., *Nucleic Acids Research* 21:5192-5197 (1993); Bubeck et al., *Nucleic Acids Research* 21:3601-3602 (1993)). In these repair mechanisms, a restriction enzyme-linearized plasmid is cotransformed with a piece of DNA with homology to either side of the gap. This method of construction requires, and is limited by, knowledge of restriction enzyme sites in the genomic DNA. Yeast and *Escherichia coli* can efficiently use this homology to repair the gap and produce a functional circular plasmid. Although these methods are effective, they are time consuming, often requiring several months of work, and require a detailed endonuclease restriction map prior to the analysis. Further, these approaches are not easily adapted to the high-throughput methodology needed today.

An alternative to the restriction analysis and subcloning methods are available where the host organism is capable of efficient homologous recombination (e.g., yeast). While most organisms require a minimum of several kilobase of homologous sequence to target integration, *Saccharomyces cerevisiae* requires more limited amounts of homology to target integration to the chromosome. Thus, a selectable marker can be amplified by PCR with targeting homology (60 bp on each side) incorporated in the PCR primers, thus avoiding the need for restriction analysis and subcloning with a large isolated genomic clone (Manivasakam et al., *Nucleic Acids Research* 23:2799-2800 (1995)). This method is the basis of an effort by the yeast researchers worldwide to disrupt all 6000+ genes in this organism. This method is practiced with normal yeast chromosomes as well as with artificial chromosomes. Treco et al. (U.S. 5,783,385) teach the identification and isolation of target DNA fragments from yeast artificial chromosome (YAC) yeast libraries, relying on the homologous recombination of transformed non-replicating plasmid DNA with the yeast genome.

Yeast provides a particularly efficient environment for homologous recombination events between genetic elements especially using linear DNA, although similar methods have been developed in bacterial organisms. For example, Zhang et al., (*Nat. Genet.* 20:123-128 (1998)) teach the site-specific

manipulation of target DNA by homologous recombination between linearized and chromosomal DNA (as well as plasmids and large episomes) in *Escherichia coli sbcA* mutants expressing RecE and RecT. These methods are useful, but require use of special mutant strains of *E. coli* that reduce the level of DNA damaging endonucleases which may digest foreign, linearized constructs.

Murphy, K. C., (*J Bacteriol.* 180:2063-2071(1998)) have reported the use of λ recombination system (Red) for generating gene knockouts in *Escherichia coli* using linearized DNA. However, use of this system for mutagenesis suffers from requirement of extra steps for transfer of λ Red functions to the host strain that needs to be mutated.

Several reports including the above mentioned have demonstrated that in the absence of specialized genetic backgrounds, integration of disrupted genes by allelic exchange does not occur in *Escherichia coli* genome.

El Karoui et al. (*Nucleic Acids Res.* 27:1296-1299 (1999), and 8th Int'l. Symp. on the Genetics of Industrial Microorganisms, 28 June 1998) report that gene replacements using linear double-stranded DNA can be achieved in wild-type *Escherichia coli* if electrocompetent cells are used.

The problem to be solved is to provide a convenient method without the limitations of known techniques to create site-directed mutants of *Escherichia coli* (or other bacteria) by allelic exchange using linear fragments of DNA to identify an altered phenotype that will define the function or effect of the mutated gene. Traditionally, the production of *Escherichia coli* mutants has been accomplished utilizing suicide vectors, which requires cloning of the gene disrupted with a selectable marker in such a vector. Winans et al. (*J. Bacteriol.* 161:1219-1221 (1985)) have shown that site-directed mutants of *Escherichia coli* strains can be obtained using linearized fragments of DNA containing a selectable marker flanked by homologous DNA. However, these examples assert the requirement of specific genetic backgrounds (e.g., *recB, recC* and/or *sbcB)*. The *recB* and *recC* mutations inactivate exonuclease, preventing it from degrading the linear DNA, whereas the *sbcB* mutation restores recombination proficiency in a strain which carries *recB* and *recC* mutations.

Although the above traditional methods permit generation of reliable site-directed mutants, there are clear disadvantages to both methods. The first one requires cloning into suicide vectors, which are available for only very limited number of bacteria. Additionally it requires an extra cloning step and is time consuming. Another disadvantage of suicide vectors in gene replacement efforts is that considerable background results due to single recombination events. The second method of Winans et al. suffers from having to use special genetic

backgrounds which, for several reasons may not be desirable when the organism to be mutated is *E. coli*, or not available at all (i.e., most bacteria other than *Escherichia coli*).

SUMMARY OF THE INVENTION

This invention provides a method for targeted DNA replacement in bacteria comprising introducing into a recombination-proficient bacteria a replacement cassette of DNA comprising a 5' recombinant region and a 3' recombinant region, each region independently at least 75% identical to a DNA target region of the bacteria. The replacement cassette may or may not additionally include a functional gene positioned between the two recombinant regions that upon integration into host restores gene functionality. The basis for screening is the phenotypic effect of a selectable marker that results from integration of the replacement cassette in the target sequence. Useful application of this method is to carry out site-directed DNA replacements for generation of desired DNA modifications in bacteria.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying Figures and Sequence Listing which form a part of this application.

Figure 1 shows a schematic diagram of *E. coli* genomic DNA region that harbors *yciK* gene, the location of *yciK* ORF with respect to restriction sites relevant to gene constructions in this invention. Location of primers used for amplifying *yciK* DNA for pyciKCK, pSCyciK1 and for knockout confirmations, is shown along with their Seq IDs.

Figure 2 shows the plasmid construct pSCyciK1. *yciK* DNA flanked with engineered *Hin*dIII and *Xho*I sites was cloned into vector pBluescript II SK (+).

Figure 3 shows a schematic diagram of *Hin*CII fragment of pLoxCat1 with relevant restriction sites used in engineering pyciKcat2 and pyciKcat3 and for confirmations of chloramphenicol resistance marker (SEQ ID NO:7) in mutants. Location and direction of *cat* ORF is shown as an arrow.

Figure 4 shows the plasmid construct pyciKcat2. The *Hin*CII fragment from *yciK* region (See Figure 1) was replaced with the 1.2 kb *Hin*CII-*Sma*I fragment of pLoxCat1. *cat* gene is in opposite orientation to *yciK* gene.

Figure 5 shows the plasmid construct pyciKcat3 engineered same as above. However, here *cat* gene is in same direction as *yciK* gene.

Figure 6 is a gel confirming *yciK* knockouts by genome-specific PCR.

Figure 7 is a gel confirming *yciK* knockouts by Southern hybridization of *Sst*II digested genomic DNA from wild-type and mutual *E. coli* probed with SEQ ID NO:18.

Figure 8 shows a schematic diagram to describe Overlap-Extension PCR to generate *tpiA* deletion.

Figure 9 shows the plasmid construct pRN106-2. A 2004 bp overlap extension PCR product with 73% of the *tpiA* structural gene deleted (Figure 8) was cloned into pCR-Blunt (Invitrogen) to yield the 5517 bp plasmid pRN106-2. The two extreme ends of *tpiA* are shown with an engineered *Sca*I site in between.

Figure 10 shows the plasmid construct pRN107-1, the source of the linearized DNA fragment used to generate the *tpiA* knockout. The 1.2 kb *Hin*CII fragment from pLoxCat1 containing the loxP-CmR-loxR cassette was inserted into the *Sca*I site of pRN106-2 to yield the 6763 bp plasmid pRN107-1. Using pRN107-1 as template and primers SEQ ID NO:1 and SEQ ID NO:4, the 3246 bp fragment containing *tpiA* flanking regions and the loxP-CmR-loxP cassette was PCR-amplified and this linear DNA fragment was used to transform KLP23 to generate *tpiA* mutants by double crossover.

Figure 11 shows the design for confirmation of *tpiA* knockouts by genome-specific PCR.

Applicant(s) have provided 19 sequences in conformity with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures – the Sequence Rules") and consistent with World Intellectual Property Organization (WIPO) Standard ST2.5 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administration Instructions).

SEQ ID NO:1 is the nucleotide sequence of DNA fragment containing *yciK* gene.

SEQ ID NO:2 is the forward primer for amplification of the *yciK* gene from the *Escherichia coli* genome.

SEQ ID NO:3 is the reverse primer for amplification of the *yciK* gene from the *Escherichia coli* genome.

SEQ ID NO:4 forward primer with *Hin*dIII site.

SEQ ID NO:5 reverse primer with *Xho*I site.

SEQ ID NO:6 is the nucleotide sequence of the yciK fragment flanked with *Hin*dIII and *Xho*I sites.

SEQ ID NO:7 is the *Hin*CII fragment of pLoxCat1.

SEQ ID NO:8 is the deduced amino acid sequence encoded by *yciK* open reading frame.

SEQ ID NO:9 is the forward primer used in a PCR reaction to confirm *yciK* knockout.

SEQ ID NO:10 is the reverse primer used in a PCR reaction to confirm *yciK* knockout.

SEQ ID NO:11 is a primer between *sbp* and *cdh*.

SEQ ID NO:12 is a primer within *tpiA* 3' region.

SEQ ID NO:13 is a primer within *tpiA* 5' region with an introduced *Sca*I site.

SEQ ID NO:14 is a primer within *yiiR* 5' region.

SEQ ID NO:15 is a primer at 5' end of *sbp*.

SEQ ID NO:16 is a primer at 3' end of *yiiR*.

SEQ ID NO:17 is expected PCR fragment when SEQ ID NO:9 and SEQ ID NO:10 were used as primers with *Escherichia coli* genomic DNA as template.

SEQ ID NO:18 is a 969bp *Sst*II fragment of pyciKCK.

SEQ ID NO:19 is the sequence of expected band in Southern experiment when *Sst*II digested wild-type genomic DNA is probed with 969bp *Sst*II fragment (SEQ ID NO:18) of pyciKCK.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a rapid and specific method for the modification or disruption of genes or other nucleic acid fragments. The thus-modified or disrupted genes may then be transformed into their native hosts and screened for an altered phenotype, indicating gene function. The invention creates stable site-directed alterations in *E. coli* or other gram negative enteric bacterial genome by allelic exchange using linear fragments of DNA to identify an altered phenotype that will define the function or effect of the mutated gene. The invention does not have the limitations of the prior art with regard to time and restriction to specific genetic backgrounds.

Additionally the present method has application in the areas of functional genomics, proteomics, metabolomics where gene disruptions to reveal function of unknown genes (and their products) or unknown function of known genes (and their products) or other functional DNA elements is desirable. Similarly the present method may be used for metabolic engineering in techniques such as gene silencing by disruption, gene integration or replacement to add genes of new function or to replace a defective gene with a functional gene. In similar fashion the present method may be useful to regulate or modify genes or operons by

altering or modifying regulatory sequences such as promoters, ribosomal binding sites, terminators, and enhancers, for example.

In this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

"Polymerase chain reaction" is abbreviated PCR.

"Open reading frame" is abbreviated ORF.

"Isopropyl β-D-Thiogalactoside" is abbreviated IPTG.

"Sodium dodecyl sulfate" is abbreviated SDS.

"Luria-Bertani" is abbreviated LB.

"Gene" refers to a nucleic acid fragment that expresses a specific RNA only or a specific RNA and protein, including regulatory sequences preceding (5' noncoding sequences) and following (3' noncoding sequences) the coding sequence. The terms "synthetic genes" and "synthetic DNA" refer to genes and DNA that can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene constructs. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled in vitro. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. "Gene constructs" may contain a full gene or less or more than a full gene.

The term "gene disruption" is used interchangeably with the term "gene knock out" to refer to the process of interfering within a gene such that no functional gene product is expressed.

The term "gene replacement" refers to a process which replaces a gene with either a functional or a mutated gene such that either no gene product is expressed or a mutant gene product is expressed.

The term "DNA replacement" means a process in which one fragment of DNA is replaced with another DNA fragment.

The term "gene modification" means any process where a gene is altered in any way including gene disruption or gene replacement.

The term "gene targeting" refers to a process where a specific site within a gene or nucleic acid fragment is identified or targeted on the basis of sequence.

The term "nucleic acid fragment" refers to a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. A nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, or synthetic DNA. A nucleic acid fragment may be a portion of a gene, may be

synthetic, or may be a genetic regulatory element. A "target nucleic acid fragment" or "target gene" or "target DNA" is any nucleic acid fragment that is inserted into a modification plasmid that is targeted for modification or disruption.

The terms "plasmid", "vector", and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for expression of that gene in a foreign host.

The term "selectable marker" refers to those genes encoding proteins that may be expressed and/or which convey a phenotype on the host that enables selection. A defective or functional gene may itself act as a selectable marker if its presence leads to a phenotype that can be selected against the background. "Background" is undesirable colonies that appear during screening of gene replacement events. The term "modification plasmid" refers to a specialized plasmid for use in the present invention comprising, at a minimum, a plasmid-specific marker and a replacement cassette.

"Enteric-specific selectable markers" are those genes encoding proteins that may be expressed in Enteric bacteria (e.g., Salmonella sp., Escherichia sp.), which convey a phenotype on the enteric host that enables selection. "Enteric bacteria" are members of the family *Enterobacteriaceae*, and include such members as *Escherichia*, *Salmonella*, and *Shigella*. They are gram-negative straight rods, 0.3-1.0 X 1.0-6.0 μm, motile by peritrichous flagella, except for *Tatumella*, or nonmotile. They grow in the presence and absence of oxygen and grow well on peptone, meat extract, and (usually) MacConkey's media. Some grow on D-glucose as the sole source of carbon, whereas others require vitamins and/or mineral(s). They are chemoorgano-trophic with respiratory and fermentative metabolism but are not halophilic. Acid and often visible gas is produced during fermentation of D-glucose, other carbohydrates, and polyhydroxyl alcohols. They are oxidase negative and, with the exception of

Shigella dysenteriae 0 group 1 and *Xenorhabdus nematophilus*, catalase positive. Nitrate is reduced to nitrite except by some strains of *Erwinia* and *Yersina*. The G + C content of DNA is 38-60 mol % (T_m, Bd). DNAs from species from species within most genera are at least 20% related to one another and to *Escherichia coli*, the type species of the family. Notable exceptions are species of *Yersina*, *Proteus*, *Providenica*, *Hafnia* and *Edwardsiella*, whose DNAs are 10-20% related to those of species from other genera. Except for *Erwinia chrysanthemi* all species tested contain the enterobacterial common antigen (Bergy's Manual of Systematic Bacteriology, D. H. Bergy, et al., Baltimore: Williams and Wilkins, 1984).

The terms "KO vector" or "knock out vector" refer to a modification plasmid lacking the target gene to be modified.

The term "knock out plasmid" refers to a modification plasmid that contains a gene targeted for disruption.

The term "knock out cassette" refers to a modification cassette that contains recombination regions designed to insert a modifying nucleic acid within the coding region of the target gene so as to prevent the effective expression of that target gene.

A "replacement cassette" refers to a specialized DNA cassette that comprises at a minimum, a selectable marker, a modifying DNA and 5' and 3' recombination regions. Replacement cassettes may optionally also comprise other modifying DNA or RNA sequences, inserted between the flanking recombination regions. Within the context of the present invention, replacement cassettes containing modifying DNA interact with the DNA to be replaced via the mechanism of homologous recombination to permit modification or disruption of the target gene or target nucleic acid.

The term "genomic host" refers to the cell or host from which the target gene or DNA has been cloned.

The term "inserting nucleic acid fragment" refers to a DNA ("inserting DNA fragment") or RNA ("inserting RNA fragment") molecule residing in a replacement cassette that is useful for the modification or disruption of a target gene or target nucleic acid fragment. The inserting nucleic acid fragment will insert at a site directed by the sequence of the recombination regions on the cassette. "Inserting DNA" may be either "modifying" or "disrupting". "Modifying DNA" or "modifying nucleic acid fragments" will result in the altering of the composition or function of a target gene but will not disrupt the gene. "Disrupting DNA" or "disrupting nucleic acid fragments" will have the effect of disrupting the target gene of interest. "Modifying DNA" or "disrupting

DNA" will include but will not be limited to non-specific DNA or RNA sequences, selectable markers, origins of replication, antisense sequences and regulatory elements. As used herein, "regulatory elements" refer to nucleotide sequences located upstream (5'), within, and/or downstream (3') to a coding sequence, which control the transcription and/or expression of the coding sequences, potentially in conjunction with the protein biosynthetic apparatus of the cell. In artificial DNA constructs regulatory elements can also control the transcription and stability of antisense RNA. One specific regulatory element is a "promoter". "Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The term "recombination region" refers to 3' and 5' flanking nucleic acid regions on the replacement cassette. Recombination regions are designed to be either complementary to, or have significant base identity with, the corresponding regions of the target genes or target nucleic acid fragments to be disrupted or modified. Significant identity, for example, may range from about 70% between the bases of the target gene ("target region") and the recombination regions. Recombination regions at the 5' end of the cassette are referred to as "5' recombination regions" and recombination regions at the 3' end of the cassette are referred to as "3' recombination regions". The maximum length of the flanking recombination region is set at approximately 150 bases when using PCR amplification in the technique.

The term "complementary" is used to describe the relationship between nucleotide bases that are hybridizable to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine.

The term "percent identity" is a relationship between two or more polynucleotide sequences as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polynucleotide

sequences as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993); Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991).

Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG Pileup program found in the GCG program package, as used in the instant invention, using the Needleman and Wunsch algorithm with their standard default values of gap creation penalty=12 and gap extension penalty=4 (Devereux et al., Nucleic Acids Res. 12:387-395 (1984)), BLASTP, BLASTN, and FASTA (Pearson et al., Proc. Natl. Acad. Sci. U.S.A. 85:2444-2448 (1988). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul et al., Natl. Cent. Biotechnol. Inf., Natl. Library Med. (NCBI NLM) NIH, Bethesda, MD 20894; Altschul et al., J. Mol. Biol. 215:403-410 (1990)). Another preferred method to determine percent identity, is by the method of DNASTAR protein alignment protocol using the Jotun-Hein algorithm (Hein et al., Methods Enzymol. 183:626-645 (1990)). Default parameters for the Jotun-Hein method for alignments are: for multiple alignments, gap penalty=11, gap length penalty=3; for pairwise alignments ktuple=6. By way of illustration, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

The term "a partial sequence" or "a portion of a sequence" refers to a sequence of sufficient length to permit homologous recombination according to

the conditions of the present method. Typically, "a partial sequence" or "a portion of a sequence" will range from about 15 bp to about 200 bp.

The term "coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' noncoding sequences), within, or downstream (3' noncoding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "operon" refers to a group of operably linked genes and regulatory elements that functions as a unit of transcription. An operon starts with a promoter, which binds RNA polymerase and initiates transcription of the operon.

The term "expression" refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

The term "recombination proficient" means that a microorganism is capable of integrating extra-chromosomal DNA into its genome via homologous recombination. Additionally, a "recombination proficient" microorganism is also capable of recombining DNA resident on chromosome if sufficient homology exists between another DNA locus.

"Transformation" refers to the transfer of a nucleic acid fragment into the host organism. "Stable transformation" refers to integrating the nucleic acid into the genome of the cell.

"Absolute recombination frequencies" range from 2.3×10^{-6} to 1.5×10^{-1} in *E. coli* depending on the system and length of homology involved. *RecA* is the most crucial component for the homologous recombination reaction, although *recA* independent mechanisms also exists. Homologs and analogs of *recA*-like proteins and other genes linked to homologous recombination have been discovered in a wide range of bacteria and eukaryotes. (Lloyd, R. G. and Low,

K. B. (1996) Homologous Recombination. In Escherichia coli and Salmonella Cellular and Molecular Biology, Ed. F.C. Neidhardt, ASM Press, p 2236-2255).

For a gene replacement to occur which requires two homologous recombination event absolute recombination frequencies can be calculated to be in the range of 10^{-12} to 10^{-2}.

Linear DNA is the replacement cassette that is isolated from a modification plasmid by restriction digestion or by PCR amplification as described in Examples 1 and 4. Preparation of linear DNA is also described in Example 3.

Although linear DNA has been used for gene replacements in some bacteria such as *Bacillus subtilis*, literature has strongly indicated that gene replacements using linear DNA do not occur in wild type bacteria such as *E. coli* and *Salmonella* unless some mutations or special genes are introduced to render these organisms hyper-recombinant. Use of electroporation, offering the highest efficiencies of transformation among the presently known methods, demonstrates that linear DNA can be used for gene replacements in wild type *E. coli*. By achieving high efficiencies of transformation, gene replacements can be performed in any recombination proficient bacteria which carry a functional recombination system.

The present method is useful for the rapid disruption or modification of genes or other genetic elements. Genes of unknown function which are disrupted in this fashion are transformed into genomic hosts where screening for an altered phenotype will reveal their function.

This method is also useful for replacing defective or mutated genes from bacterial genome with the functional gene provided a selection phenotype exist for the function of the gene.

The gene replacement techniques described in examples to follow can also be used for regulation of genes present in operons by replacement of operably linked promoter region with a promoter of desired strength.

The method described in this invention to obtain gene replacement can be used in enteric bacteria without specialized host backgrounds or use of suicide vectors. Electrotransformation also constitutes a clear-cut method to obtain gene replacements with linear DNA in wild-type *Escherichia coli* on plasmid and chromosomal targets. However, any methods (chemical transformation or electrotransformation) that ensure a transformation efficiency high enough to enable the occurrence of double homologous recombination between the target DNA and the replacement cassette will work, in this specific case electro-transformation was used. The method of electrotransformation itself does not

guarantee a success in gene replacements if quality of competent cells is not good or if host strains deficient in recombination functions are used. A high efficiency of transformation must occur in recombination proficient bacteria. Depending on the absolute recombination frequencies of target gene (or target DNA) locus and the host strain, required transformation efficiencies will vary. Use of long PCR primers as described in Example 3 completely eliminates any cloning requirements for gene replacements.

Therefore, the present method is useful for the rapid disruption or modification of genes or other genetic elements. Genes of unknown function disrupted in this fashion may be transformed into genomic hosts where screening for an altered phenotype will reveal their function.

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLES

GENERAL METHODS

Procedures required for PCR amplification, DNA modifications by endo- and exonucleases for generating desired ends for cloning of DNA, ligations, and bacterial transformation are well known in the art. Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989) (hereinafter "Sambrook"); and by T. J. Silhavy, M. L. Bennan, and L. W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring, N.Y. (1984) and by Ausubel, F. M. et al., Current Protocols in Molecular Biology, John Wiley & Sons, Inc. (1994-98).

For computational nucleic acid and protein sequence analysis Wisconsin Package Version 9.0 and 10.0, Genetics Computer Group (GCG) and Vector NTI Deluxe v4.0.3 software and database packages were used. Unless specified otherwise, all default paramenters were used.

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in Manual of Methods for General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W.

Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds.), American Society for Microbiology, Washington, DC. (1994)) or by Thomas D. Brock, in Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

PCR reactions were run on GeneAMP PCR System 9700 using Amplitaq or Amplitaq Gold enzymes (PE Applied Biosystems, Foster City, CA). The cycling conditions and reactions were standardized according to manufacture's instructions.

The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "μL" means microliter, "mL" means milliliters, "L" means liters, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "Cm^r" means chloramphenicol resistance, "cat" means chloroamphenicol acetyl transferase, "Amp^r" means ampicillin resistance, "Amp^s" means ampicillin sensitivity, "kb" means kilo base.

EXAMPLE 1

Targeted Gene Replacement by Linear DNA Fragments for Site-directed Mutagenesis of *Escherichia coli*

Example 1 describes the cloning, mutation and integration of a mutant *yciK* gene into the *Escherichia coli* genome by homologous recombination using linearized mutant DNA.

Isolation of *yciK* gene:

A 1460 bp fragment of DNA (SEQ ID NO:1, Figure 1) containing *yciK* gene was amplified using pCR-Script Amp Cloning kit from Stratagene (catalog #211188) from genomic DNA of *Escherichia coli* strain W3110 (Sigma; catalog #D0421). PCR amplification (Mullis and Faloona, *Methods Enzymol.* 155:335-350 (1987)) was performed using oligonucleotides SEQ ID NO:2 and SEQ ID NO:3 as forward- and reverse primers, respectively (Figure 1). The PCR product was blunt-ended and ligated into *Srf*I site of cloning vector pCR-Script Amp SK(+) followed by transformation into XL-1 Blue MRF' Kan supercompetent cells. Transformed cells were plated on Luria-Bertani medium containing ampicillin, X-gal and IPTG. Plasmids were isolated from white, ampicillin-resistant transformants and tested for the presence of the insert by restriction digestion. One such plasmid, pyciKCK, was subsequently confirmed

for the insert containing *yciK* gene by sequencing on ABI PRISM (Model 377, Version 2.1.1).

Cloning of *yciK* gene in a vector with suitable restriction sites:

The pyciKCK DNA was used as a template with primers SEQ ID NO:4 and SEQ ID NO:5 to generate a DNA fragment (SEQ ID NO:6) containing *yciK* gene flanked with *Hin*dIII and *Xho*I restriction sites (Figure 1) for cloning in *Hin*dIII/*Xho*I restricted pBluescript II SK(+) vector (Stratagene; catalog #212205). The resultant plamid was named pSCyciK1 (Figure 2).

In vitro disruption of *yciK* gene with insertional mutagenesis:

To disrupt *yciK* open reading frame (ORF), a *Hin*cII fragment (Figure 2) of the plasmid pSCyciK1 which encodes amino acids 124 to 189 of *yciK* ORF (total length=252 amino acids; SEQ ID NO:8) was replaced with a 1.2 kb *Hin*CII-*Sma*I fragment (bases 1-1225 of the SEQ ID NO:7 (see also Figure 3), constituting the sense strand of the selectable chloramphenicol-resistance encoding gene flanked by bacteriophage P1 *lox*P sites (Snaith et al., *Gene* 166:173-174 (1995) in plasmid pLoxCat1). This construction yielded two plasmids pyciKcat2 (Figure 4) and pyciKcat3 (Figure 5) containing Cm^r ORF (cat) in opposite or same orientation as *yciK*, respectively.

Introduction of *yciK* mutation into *Escherichia coli* genome by homologous recombination using linearized mutant DNA:

The plasmid construct (pyciKcat2 or pyciKcat3) was digested with *Afl*III and *Xmn*I, followed by gel purification of a 3.9 kb band of linearized DNA fragments (Figures 3 and 4) containing mutant *yciK* gene. *Escherichia coli* strains FM5 (ATCC 53911) and KLP23 (FM5 *glpK*$^-$*gldA*$^-$) were electrotransformed with 0.5 μg of the above 3.9 kb linear DNA and the resulting transformants were screened for double recombinant phenotype of chloramphenicol resistance (25 μg/mL) and ampicillin sensitivity (100 μg/mL) on LB medium. High numbers of transformants (roughly 10^7-10^9) were screened. Genomic DNA from Cm^rAmp^s colonies was dot blotted on Hybond N^+ positively charged nylon membrane (Amersham; catalog # RPN2020B). A 982 bp *Xba*I fragment (bases 58-1039 of SEQ ID NO:7 denote the sense strand) that contains Cm^r encoding fragment, was used a probe. All Cm^rAmp^s colonies tested positive for the presence of DNA encoding chloramphenicol resistance gene under high stringency conditions (0.1 x SSC, 0.1% SDS, 60°C for 15 min). Probe labeling, hybridization and detection for dot blots and subsequent Southern blotting experiments, were performed using ECL random prime labelling and detection systems version II, (Amersham International plc, Buckinghamshire, England). Genomic DNA from randomly selected Cm^rAmp^s colonies, that also tested

positive for the presence of the *cat* (encodes chloramphenicol acetyl transferase) gene, was used as template for PCR reactions using oligonucleotide pair (SEQ ID NO:9 and SEQ ID NO:10; Figure 1). These primers were designed such that only genome-specific *yciK* (wild-type or mutant) amplification could occur. Plasmid-borne *yciK* (wild-type in pSCyciK1 or mutant in pyciKcat2/pyciKcat3) did not contain complementary sequences to these primers and as a result could not yield any PCR product. This design was for avoiding false positives. The DNA from mutant colonies yielded a 3.1 kb PCR product compared to 2.1 kb product from wild-type (SEQ ID NO: 17; Figure 6) confirming that *yciK* gene in the *Escherichia coli* genome contained an insertion as predicted by the construct design. The mutants were further authenticated by Southern hybridization of *Sst*II digested genomic DNA fragments (Figure 1) probed with 969 bp *Sst*II fragment of pyciKCK containing *yciK* DNA (SEQ ID NO:18). All the mutants yielded a 2.3 kb band compared to the 1.3 kb band (SEQ ID NO:19) of wild-type strains (Figure 7).

This Example demonstrated that several independent *yciK* knockouts were obtained using two different gene constructs and two different genetic backgrounds of *E. coli* hosts. FM204 was a FM5 derivative obtained by integrating linearized DNA from pyciKcat2; the FM5 integrants that resulted from the pyciKcat3 *Afl*III-*Xmn*I fragment were named FM301 and FM302. Similarly, KLP201 and KLP202 were KLP23 derivatives using pyciKcat2 fragments. KLP301 and KLP304 were also derived from KLP23 by using the fragment from pyciKcat3.

EXAMPLE 2

Targeted Replacement of Mutant Gene with a Functional Gene Using Linear DNA Fragments

Linear DNA molecules can also be used to replace mutant genes with the wild-type gene. If a mutant has a phenotype of lack of growth under certain nutrient or environmental conditions, a replacement of wild-type gene can be selected by acquired ability for growth under same conditions. For example, *hemB* mutants of *Escherichia coli* and *Bradyrhizobium japonicum* require exogenously added hemin for growth (Chauhan, S. and O'Brian, M. R. (*J. Bacteriol.* 175:7222-7227 (1993)). A linear DNA fragment containing functional *hemB* gene from a homologous or heterologous source is amplified (by PCR or by cloning). The *Escherichia coli hemB* mutant strain RP523 (Li, J. M. et al., *J. Bacteriol*, 170:1021-1025 (1988) is electrotransformed with the amplified linear(ized) DNA fragments that harbor functional *hemB* gene. Transformants are screened on LB plates for growth. Any colonies that appear are due to a

functional *hemB* gene in the genome. If insertional mutations are being replaced, genomic constructs are further tested by PCR amplification or Southern hybridizations of the region of DNA modified. If the insertion also contains a selectable marker, the transformants may be screened for the absence of the phenotype conferred by the selectable marker. In case a point mutation is to be replaced, the PCR amplified fragment is sequenced either directly or after cloning in a suitable vector (e.g. pBluescript II SK(+), Stratagene, Cat #212205) for the presence of wild-type sequence. This method is extremely beneficial for restoring specific gene functionality in a host strain that has gone through several sequential mutations in the genome.

EXAMPLE 3

One-step PCR Method to Generate Linear DNA Fragments for Site-directed Mutagenesis

Oligonucleotide pairs approximately 150 bases in length are synthesized. Up to 30 bp on the 3'-end of forward primer is identical to a region flanking the 5'-end of a selectable marker and the rest of the bases are identical to the 5' region of the gene to be disrupted. Up to 30 bp on the 3'-end of reverse primer are complementary to a region flanking the 3'-end of a selectable marker and the rest of the bases are identical to the 3' region of the gene to be disrupted. The above oligonucleotides are used in a PCR reaction to amplify DNA fragments that contain a marker gene flanked on each side with at least 20 bases of sequence homologous to the gene desired to be mutated. The PCR product is used to electrotransform wild-type *Escherichia coli* cells for generating mutants by homologous recombination which can be tested by methods described in Example 1.

EXAMPLE 4

Engineering of Triosephosphate Isomerase Mutant of *Escherichia coli* KLP23

This example describes the construction of plasmid for triosephosphate isomerase gene replacement in *Escherichia coli* KLP23 and engineering of triosephosphate isomerase mutant RJ8m by linear DNA transformation.

Construction of plasmid for triosephosphate isomerase gene replacement in *Escherichia coli* KLP23:

Escherichia coli KLP23 genomic DNA was prepared using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN). A 1.0 kb DNA fragment containing *cdh* and the 3' end of triosephosphate isomerase (*tpiA*) genes was amplified by PCR (Mullis and Faloona, *Methods Enzymol.* 155:335-350 (1987)) from KLP23 genomic DNA using primers (SEQ ID NO:11 and SEQ ID NO:12). A 1.0 kb DNA fragment containing the 5' end of *tpiA*, *yiiQ*, and the 5' end of *yiiR*

genes was amplified by PCR from KLP23 genomic DNA using primers SEQ ID NO:13 and SEQ ID NO:14. The first 26 bases at 5' end of primer SEQ ID NO:13 are complementary to the primer SEQ ID NO:12 to enable subsequent overlap extension PCR, followed by a *Sca*I (AGTACT) site. The next 28 bases following *Sca*I site at the 3' end of SEQ ID:13 reside on 5' end of the tpiA gene (Figure 8). The gene splicing by overlap extension technique (Horton et al., *BioTechniques* 8:528-535 (1990)) was used to generate a 2.0 kb fragment by PCR using the above two PCR fragments as templates and primers SEQ ID NO:11 and SEQ ID NO:14. This fragment represented a deletion of 73% of the 768 bp *tpiA* ORF. Overall, this fragment had 1.0 kb flanking regions on either side of the *Sca*I cloning site (within the partial *tpiA*) to allow for chromosomal gene replacement by homologous recombination.

The above blunt-ended 2.0 kb PCR fragment was cloned into the pCR-Blunt vector using the Zero Blunt PCR Cloning Kit (Invitrogen, San Diego, CA) to yield the 5.5 kb plasmid pRN106-2 (Figure 9) containing kanamycin and Zeocin resistance genes. The 1.2 kb *Hinc*II fragment (SEQ ID NO:7, Figure 3) containing a chloramphenicol-resistance gene flanked by bacteriophage P1 *lox*P sites (Snaith et al., Gene 166:173-174 (1995)), was used to interrupt the *tpiA* fragment in plasmid pRN106-2 by ligating it to *Sca*I-digested plasmid pRN106-2 to yield the 6.8 kb plasmid pRN107-1 (Figure 10).

Engineering of triosephosphate isomerase mutant (RJ8m) by linear DNA transformation:

Using pRN107-1 as template and primers SEQ ID NO:11 and SEQ ID NO:14, the 3.2 kb fragment containing *tpiA* flanking regions and chloramphenical resistance marker (SEQ ID NO:7) was PCR-amplified and gel-extracted. *Escherichia coli* KLP23 was electrotransformed with about 1 μg of this 3.2 kb linear DNA fragment. Transformants that were chloramphenicol-resistant (12.5 μg/mL) and kanamycin-sensitive (30 μg/mL) were further screened on M9 minimal media for poor glucose utilization on 1 mM glucose, a phenotype expected for *tpiA* mutant. An *Eco*RI digest of genomic DNA from one such mutant, RJ8m, was probed with the intact *tpiA* gene via Southern analysis (Southern, *J. Mol. Biol.* 98:503-517 (1975)). The results indicated that it was a double-crossover integrant (*tpiA* gene replacement) as the two expected 6.6 kb and 3.0 kb bands were observed, owing to the presence of an additional *Eco*RI site within the chloramphenicol resistance gene. As expected, the host KLP23 and wild-type FM5 controls yielded single 8.9 kb and 9.4 kb bands respectively. This *tpiA* mutant was further analyzed by genomic PCR using primers SEQ ID NO:15 and SEQ ID NO:16 (Figure 11), which yielded the expected 4.6 kb PCR fragment

while for the same primer pair the host KLP23 and wild-type FM5 strains both yielded the expected 3.9 kb PCR fragment. When cell-free extracts from *tpiA* mutant RJ8m and host KLP23 were tested for *tpiA* activity using glyceraldehyde 3-phosphate as substrate, no activity was observed with RJ8m. Taken together, a confirmation of *tpiA* knockout was demonstrated.

CLAIMS

1. A method for gene replacement in recombination-proficient bacteria comprising:

(a) transforming recombination-proficient bacteria with a replacement cassette containing 5' and 3' flanking sequences, each flanking sequence independently with at least 75% identity to the target DNA; and

(b) screening for an altered phenotype.

2. The method of Claim 1 in which absolute recombination frequencies are at least 10^{-4} to enable selection of double homologous recombination event(s).

3. The method of Claim 1 in which the transforming step is performed chemically or by electroporation.

4. The method of Claim 1 in which the recombination-proficient bacteria are enteric bacteria.

5. A method of gene disruptions in recombination-proficient enteric bacteria comprising transforming the recombination-proficient enteric bacteria with a replacement cassette containing 5' and 3' flanking sequences, each flanking sequence independently with at least 75% identity to the target DNA.

6. A method of replacing mutant genes with functional genes in recombination-proficient enteric bacteria comprising transforming recombination-proficient bacteria with a replacement cassette containing 5' and 3' flanking sequences, each flanking sequence independently with at least 75% identity to the target DNA.

7. A method of regulating complex operons comprising deleting or replacing one or more promoter regions operably linked to the complex operon with at least one alternate desired promoter region.

8. A method according to any one of Claims 1, 2, 3, 4, or 5 for gene replacement in recombination-proficient bacteria further comprising: prior to transforming the recombination-proficient the replacement cassette is generated using direct PCR of a selection marker wherein the direct PCR uses long primer pairs, each primer having at least 75% identity to the target DNA at the 5' end and at least 75% identity to the selection marker at the 3' end.

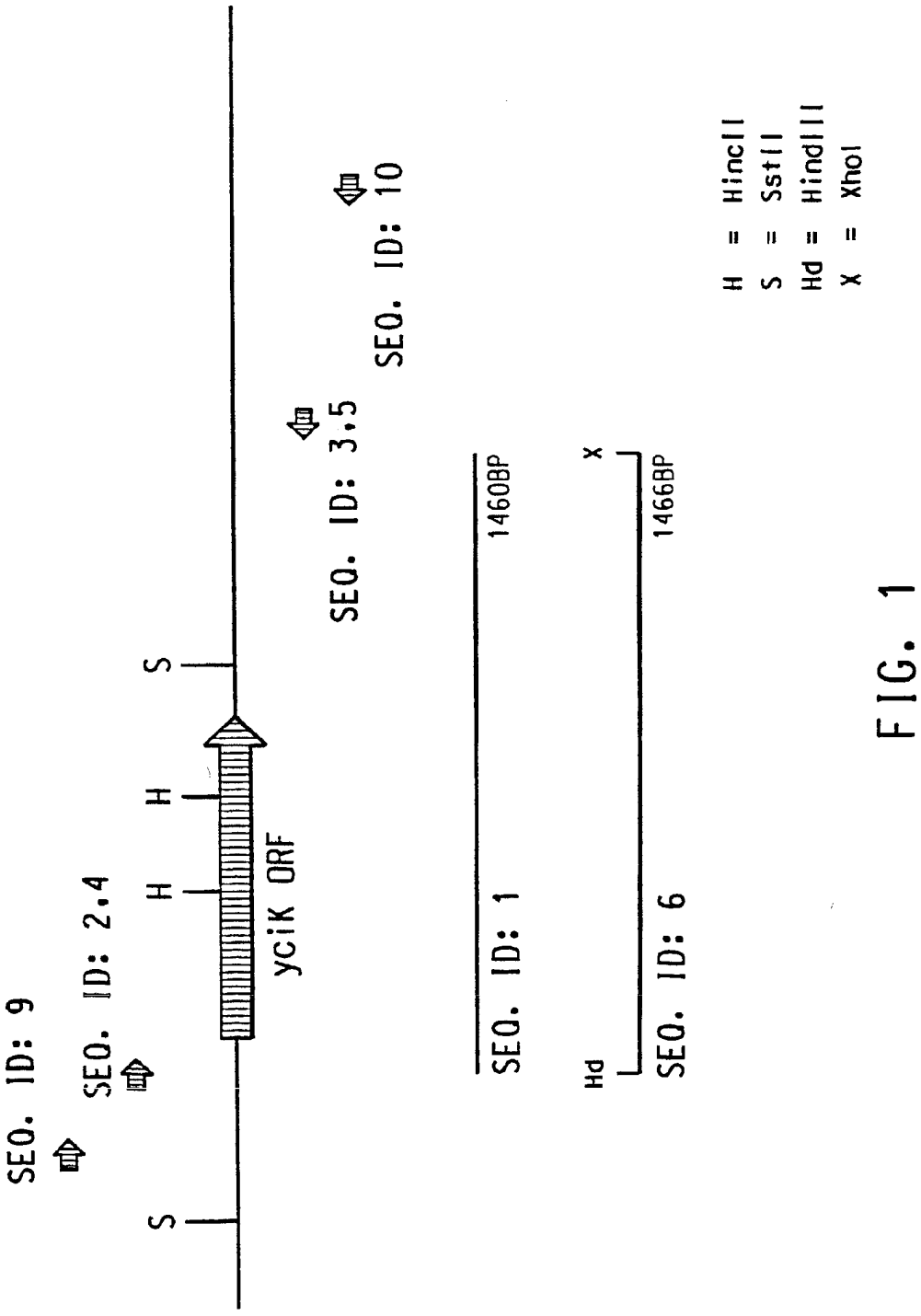

SEQ. ID: 9
SEQ. ID: 2.4
S
H
H
S
ycik ORF
SEQ. ID: 3.5
SEQ. ID: 10
1460BP
SEQ. ID: 1
X
Hd
1466BP
SEQ. ID: 6
H = HincII
S = SstII
Hd = HindIII
X = XhoI

FIG. 1

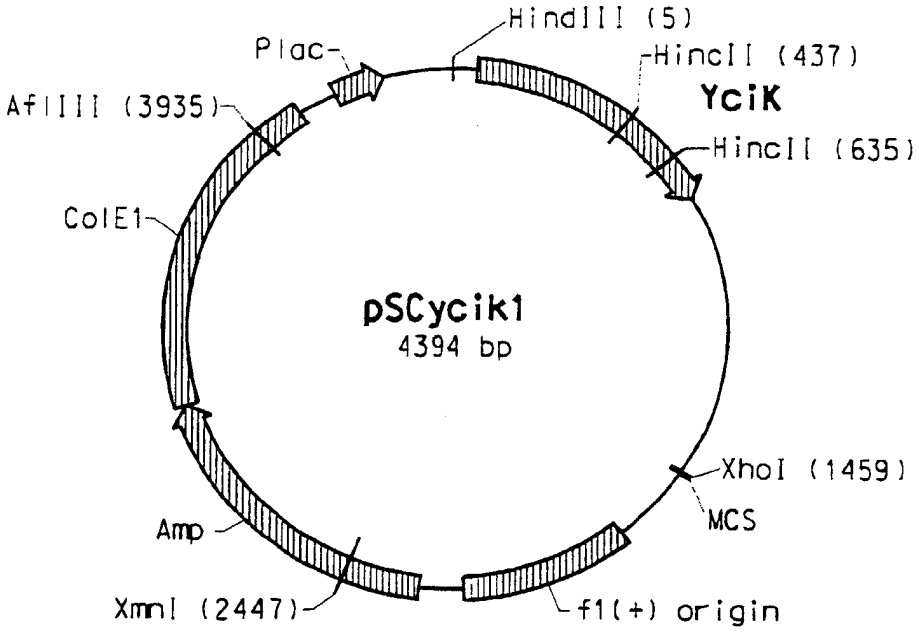


FIG. 2

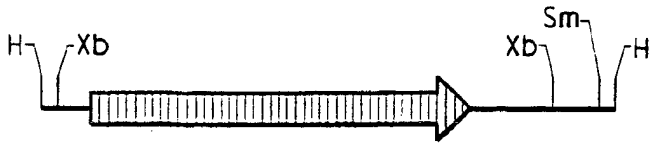


FIG. 3

H = HincIII
Xb = XbaI
Sm = SmaI

FIG. 4

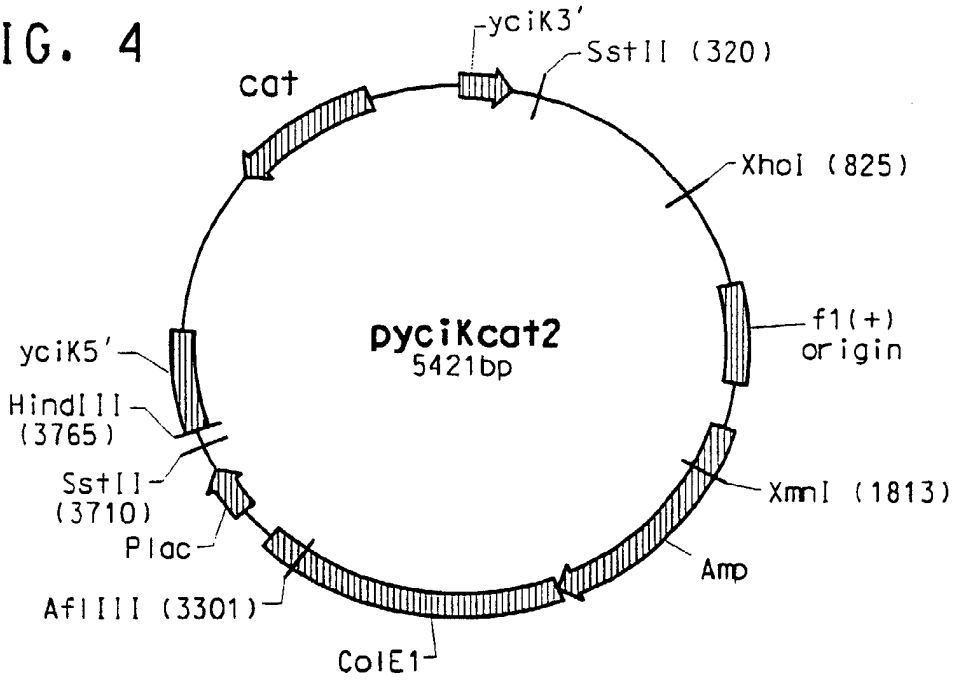
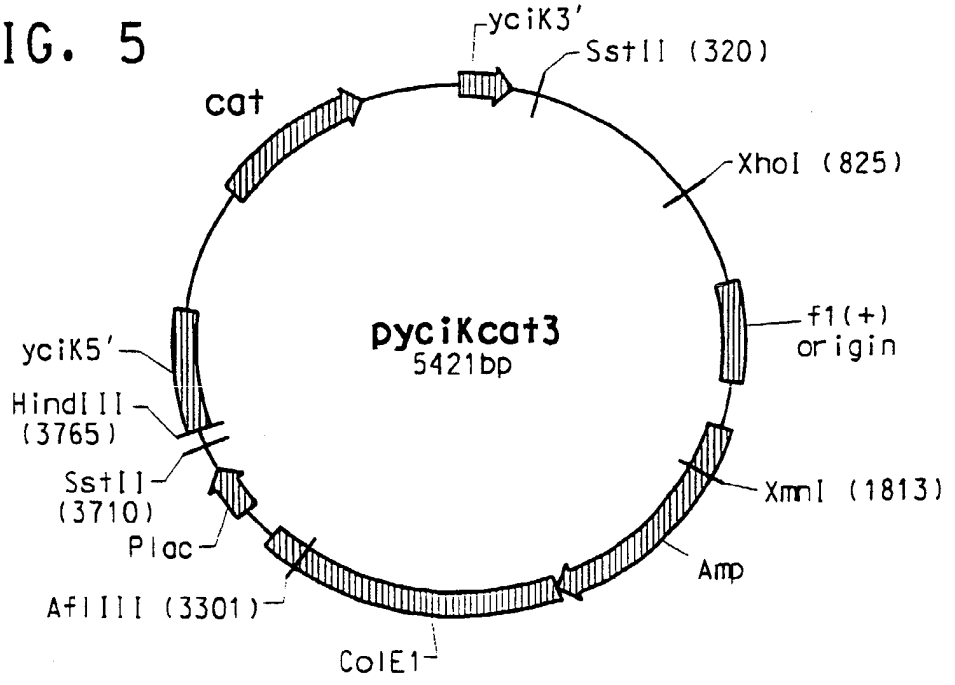

yciK3'
SstII (320)
cat
XhoI (825)
f1(+)
origin
pyciKcat2
5421bp
yciK5'
HindIII
(3765)
SstII
(3710)
Plac
AflII (3301)
XmnI (1813)
Amp
ColE1

FIG. 5


yciK3'
SstII (320)
cat
XhoI (825)
f1(+)
origin
pyciKcat3
5421bp
yciK5'
HindIII
(3765)
SstII
(3710)
Plac
AflII (3301)
XmnI (1813)
Amp
ColE1

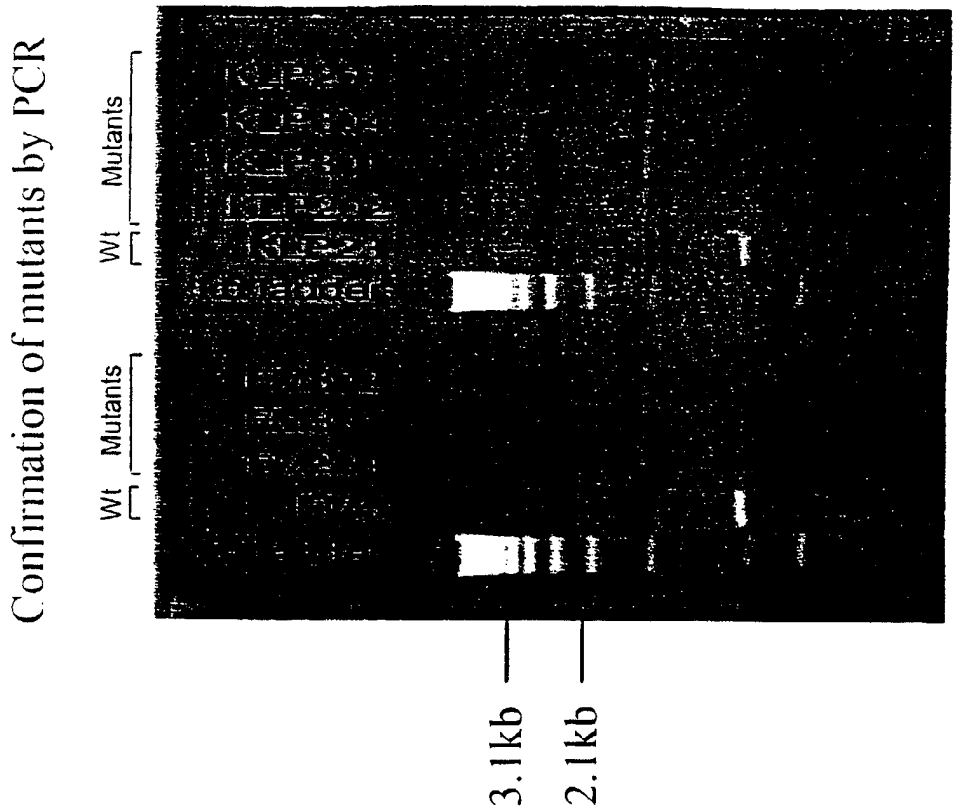

Confirmation of mutants by PCR
Wt
Mutants
Wt
Mutants
3.1kb
2.1kb

FIG. 6

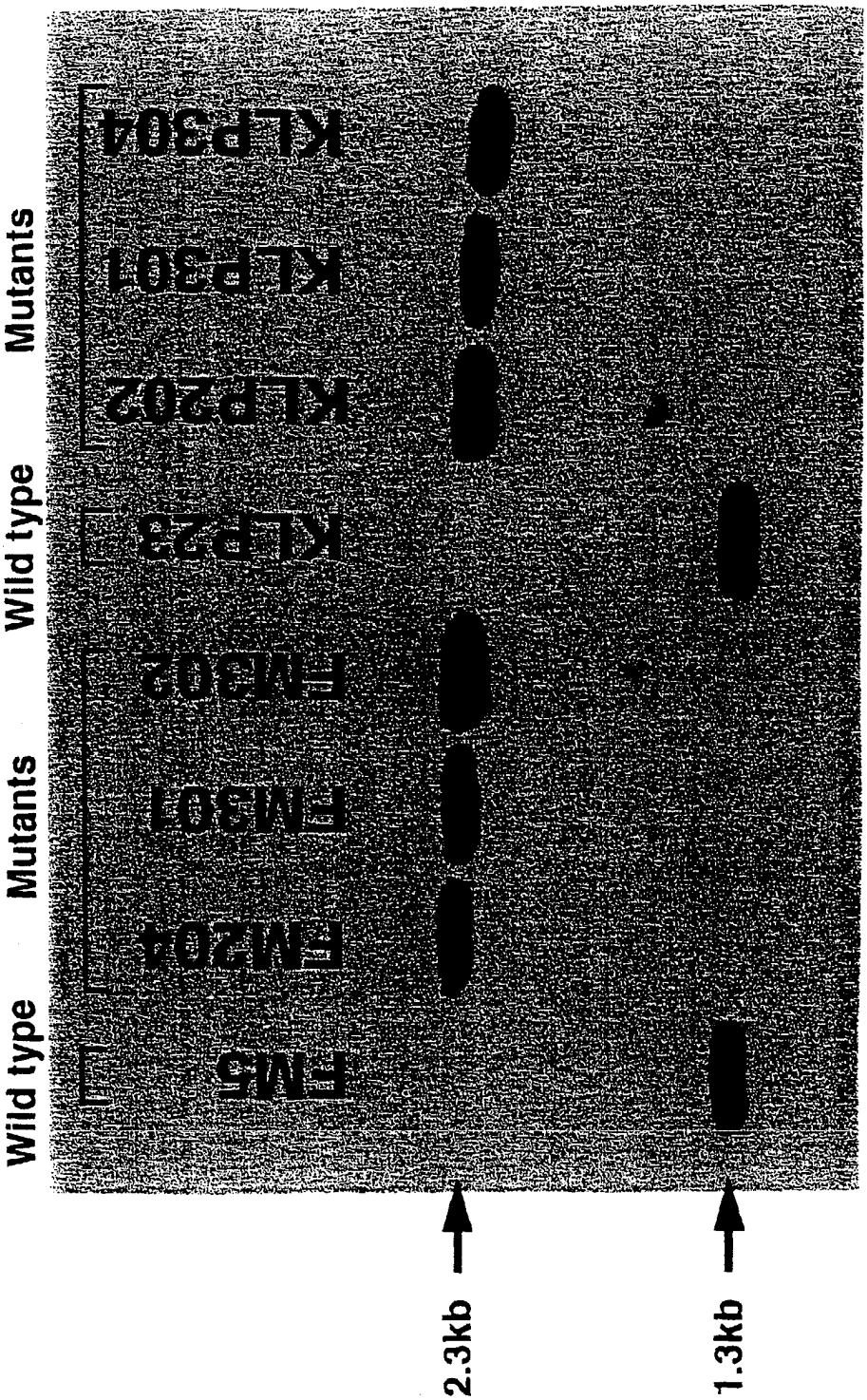

Wild type
Mutants
Wild type
Mutants
FM5
FM204
FM301
FM302
KLP23
KLP202
KLP301
KLP304
2.3kb
1.3kb

FIG. 7

Step 1:

Generation of 1 kb template fragments having 3' complementary ends using primer pairs SEQ ID NO: 11 - SEQ ID NO: 12 and SEQ ID NO: 13 - SEQ ID NO: 14. The 5' end of SEQ ID NO: 13 is complementary to SEQ ID NO: 12. A ScaI restriction site was engineered into SEQ ID NO: 13.

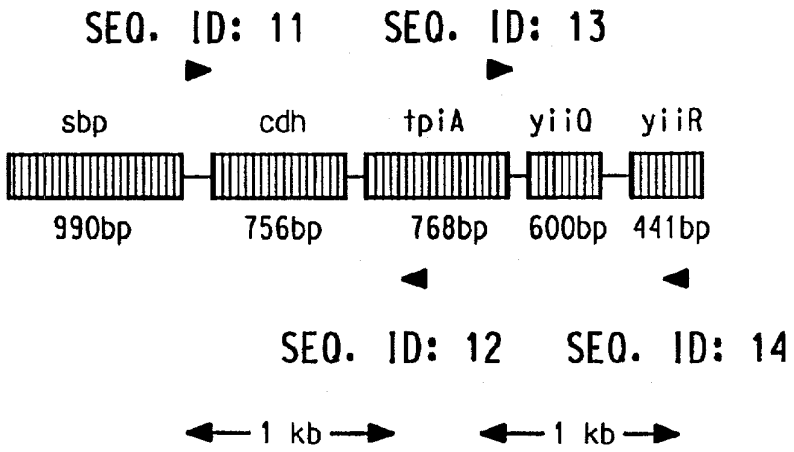


FIG. 8-1

Step 2:

Overlap-Extension PCR generation of 2 kb DNA fragment using above i kb template fragments rsulted from annealing of 3' complementary ends followed by 5' to 3' extension of both DNA strands to yield the final 2 kb DNA product. Further rounds of PCR with this 2 kb template and primers SEQ D NO: 11 and SEQ ID NO: 14 resulted in amplification of this 2 kb DNA fragment that represented a deletion of 73% of the tpiA structural gene with a unique ScaI cloning site within.

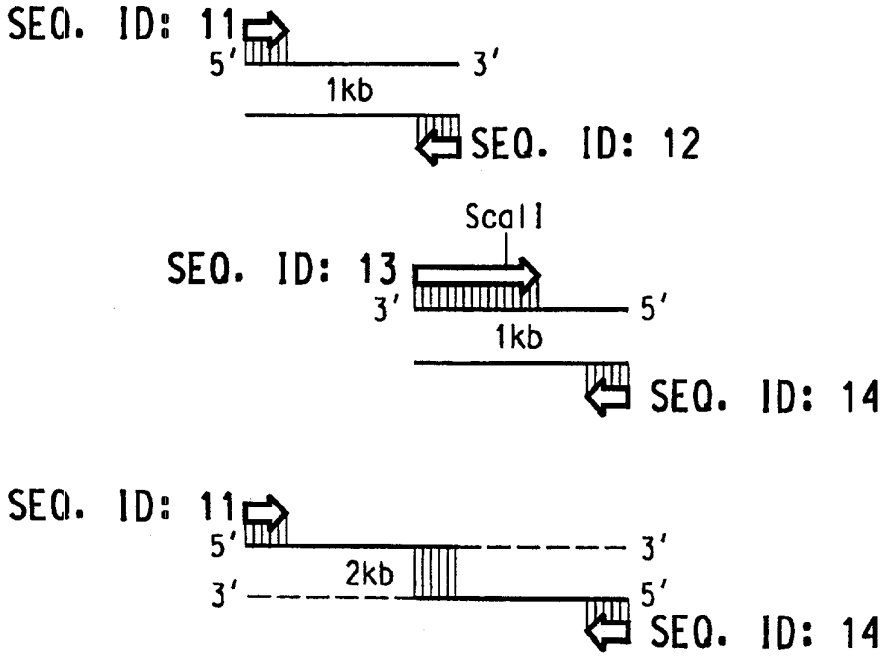


FIG. 8-2

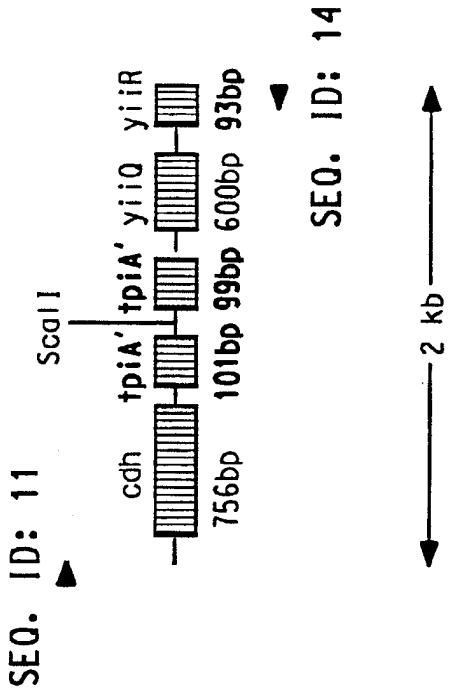

SEQ. ID: 11
ScaII
cdh
tpiA'
tpiA'
yiiQ
yiiR
756bp
101bp
99bp
600bp
93bp
SEQ. ID: 14
2 kb

FIG. 8-3

FIG. 9

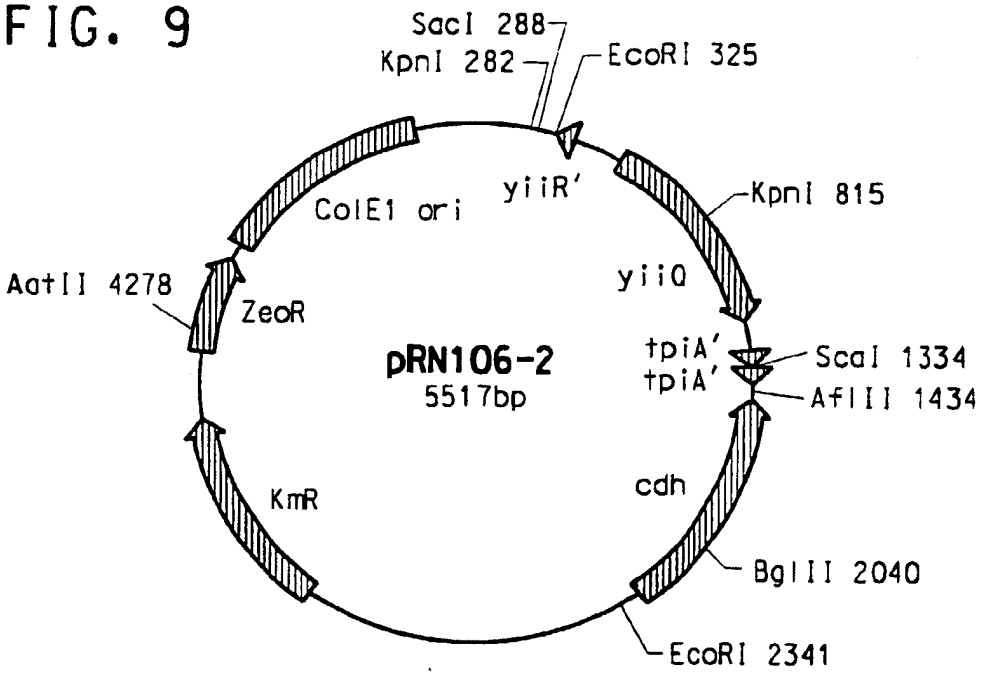
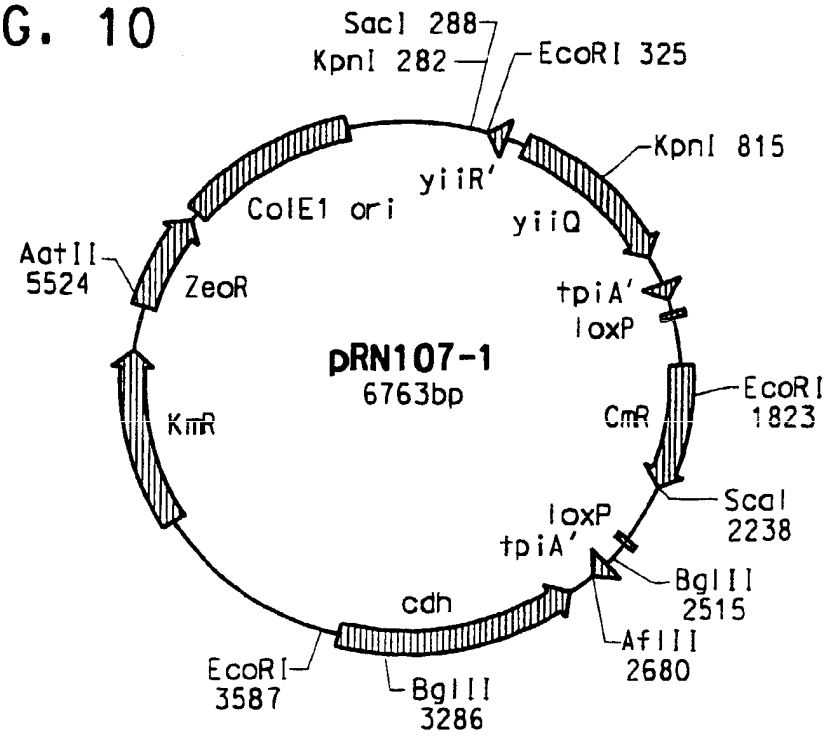

SacI 288
KpnI 282
EcoRI 325
yiiR'
KpnI 815
ColE1 ori
yiiQ
AatII 4278
ZeoR
pRN106-2
5517bp
tpiA'
tpiA'
ScaI 1334
AflII 1434
cdh
KmR
BglII 2040
EcoRI 2341

FIG. 10


SacI 288
KpnI 282
EcoRI 325
yiiR'
KpnI 815
ColE1 ori
yiiQ
AatII
5524
ZeoR
tpiA'
loxP
pRN107-1
6763bp
EcoRI
1823
KmR
CmR
ScaI
2238
loxP
tpiA'
BglII
2515
cdh
AflII
2680
EcoRI
3587
BglII
3286

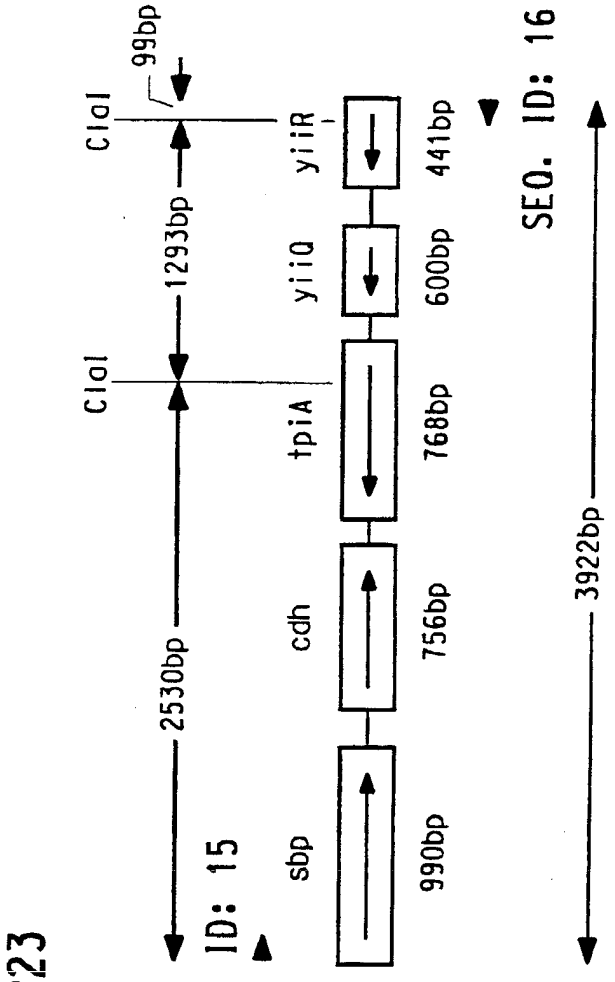

KLP23
SEQ. ID: 15
SEQ. ID: 16
ClaI
ClaI
99bp
1293bp
2530bp
3922bp
sbp
cdh
tpiA
yiiQ
yiiR
990bp
756bp
768bp
600bp
441bp

FIG. 11-1

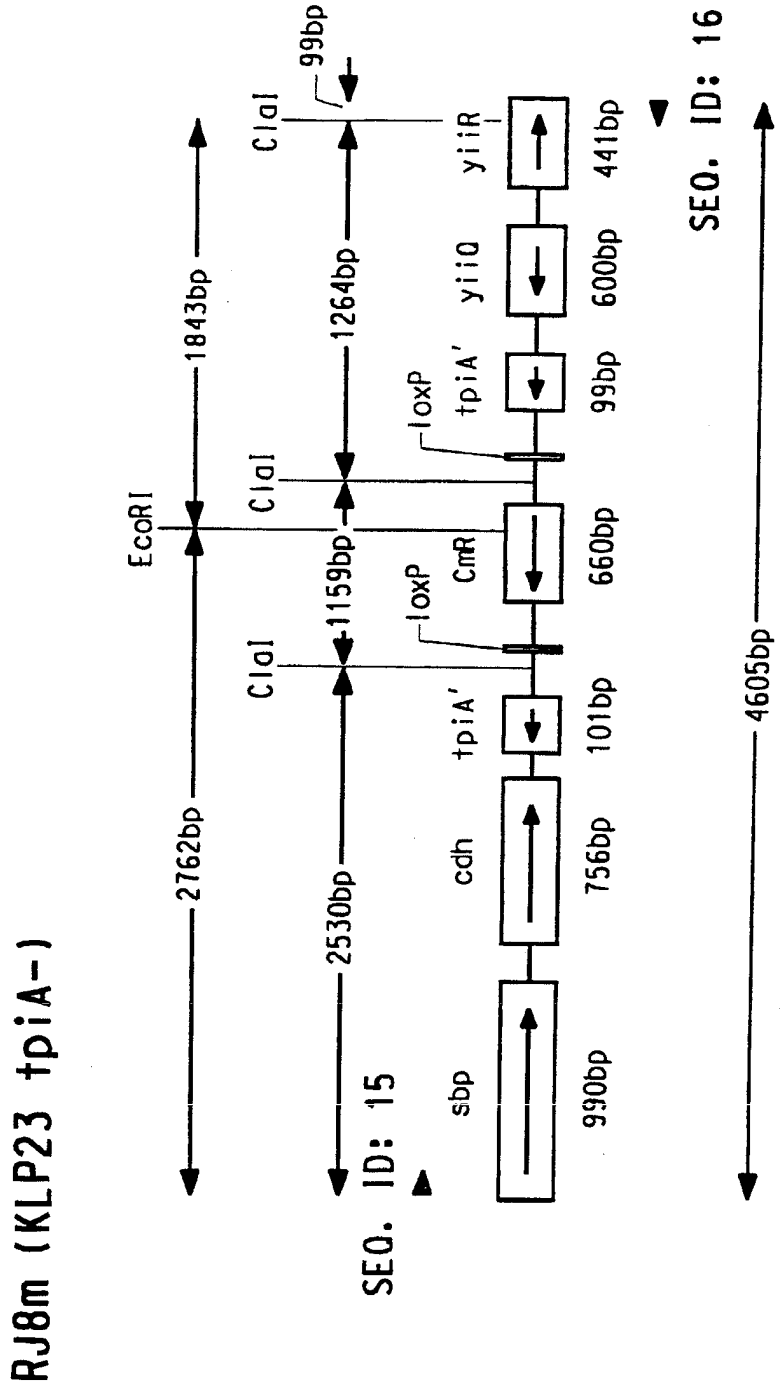

RJ8m (KLP23 tpiA-)
2762bp
1843bp
EcoRI
ClaI
2530bp
1264bp
ClaI
99bp
ClaI
1159bp
SEQ. ID: 15
sbp
990bp
cdh
756bp
tpiA'
101bp
loxP
CmR
660bp
loxP
tpiA'
99bp
yiiQ
600bp
yiiR
441bp
4605bp
SEQ. ID: 16

FIG. 11-2

SEQUENCE LISTING

<110> E. I. du Pont de Nemours and Company

<120> Targeted Gene Replacements in Enteric Bacteria Using Linear DNA

<130> BC1010 PCT

<140>
<141>

<150> 60/141,103
<151> 1999-06-25

<160> 19

<170> Microsoft Office 97

<210> 1
<211> 1460
<212> DNA
<213> Escherichia coli

<400> 1
cataaagccc caaagtcata aagtacactg gcagcgcgac agacaacgtc gttgaggagt 60
agccatgcat taccagccaa aacaagattt actcaatgat cgcattatcc tggtgacggg 120
agccagcgat ggtattggtc gtgaagccgc gatgacgtat gcacgctatg gtgcgacagt 180
gattctgttg ggccgtaatg aagaaaaatt acgtcaggta gccagccaca taaacgaaga 240
aactgggcgt cagccacagt ggtttattct cgatttgctg acctgcacgt ccgaaaattg 300
ccaacaactg gcacagcgca ttgccgttaa ttatccgcgt ctggatggtg ttttgcataa 360
tgccggattg ctcggcgatg tttgcccaat gagcgaacaa aatccgcagg tctggcagga 420
cgtcatgcag gtcaacgtta atgccacctt tatgctcacc caggcactgc ttcctttatt 480
actcaaatcg gacgccggtt cactggtctt tacttcatca agcgttggac gtcagggacg 540
agccaactgg ggtgcatatg cagcgtcgaa atttgccacc gaagggatga tgcaggtact 600
ggccgatgaa tatcagcagc gcctgcgtgt caactgcatt aacccaggcg gtacgcgcac 660
cgcaatgcgt gccagcgcct tcccgaccga agatccacag aaacttaaaa cacccgctga 720
tatcatgccg ctctacctct ggctgatggg cgatgacagc cgccgtaaaa ccggcatgac 780
ctttgacgcc caaccgggcc gtaaaccagg aatttcccaa tgagtgatga acgctaccaa 840
cagcgtcagc agcgagtgaa agaaaaagta gatgctcgtg tggcccaggc ccaggatgaa 900
cgcggtatta tcatcgtctt taccggcaat ggaaaaggca aaaccaccgc ggcatttggt 960
acggcaacac gcgcagttgg tcacggaaaa aaagtaggcg tcgtgcagtt tattaaaggc 1020
acctggccta atggcgaacg caatctgctg gagccacatg gcgttgagtt tcaggtgatg 1080
gcaacgggct ttacctggga tacacaaaac cgcgagtctg ataccgccgc ctgccgcgaa 1140
gtctggcaac atgcaaagcg gatgcttgct gattcctcac tggatatggt tttgcttgat 1200
gaactgacgt atatggtggc gtatgactat ttgccactgg aagaagtggt gcaggcgtta 1260
aatgaacgtc cacatcaaca gacggtgatt atcacgggtc gtggttgtca tcgggatatt 1320
cttgaactgg cagatacggt aagtgaatta cgccccgtca aacatgcgtt tgatgccggt 1380
gtaaaagcgc agatagggat cgattattaa tctgtaaaaa tgccggatat gacgtccggc 1440
atactttgat taaccgttat 1460

<210> 2
<211> 30
<212> DNA
<213> Primer

<400> 2
cataaagccc caaagtcata aagtacactg 30

<210> 3
<211> 29
<212> DNA
<213> Primer

<400> 3
ataacggtta atcaaagtat gccggagtc 29

<210> 4
<211> 33
<212> DNA
<213> primer

<400> 4
aggaagcttg ccccaaagtc ataaagtaca ctg 33

<210> 5
<211> 33
<212> DNA
<213> primer

<400> 5
atactcgagg ttaatcaaag tatgccggac gtc 33

<210> 6
<211> 1466
<212> DNA
<213> Escherichia coli

<400> 6
aggaagcttg ccccaaagtc ataaagtaca ctggcagcgc gacagacaac gtcgttgagg 60
agtagccatg cattaccagc caaaacaaga tttactcaat gatcgcatta tcctggtgac 120
gggagccagc gatggtattg gtcgtgaagc cgcgatgacg tatgcacgct atggtgcgac 180
agtgattctg ttgggccgta atgaagaaaa attacgtcag gtagccagcc acataaacga 240
agaaactggg cgtcagccac agtggtttat tctcgatttg ctgacctgca cgtccgaaaa 300
ttgccaacaa ctggcacagc gcattgccgt taattatccg cgtctggatg gtgttttgca 360
taatgccgga ttgctcggcg atgtttgccc aatgagcgaa caaaatccgc aggtctggca 420
ggacgtcatg caggtcaacg ttaatgccac ctttatgctc acccaggcac tgcttccttt 480
attactcaaa tcggacgccg gttcactggt ctttacttca tcaagcgttg gacgtcaggg 540
acgagccaac tggggtgcat atgcagcgtc gaaatttgcc accgaaggga tgatgcaggt 600
actggccgat gaatatcagc agcgcctgcg tgtcaactgc attaacccag gcggtacgcg 660
caccgcaatg cgtgccagcg ccttcccgac cgaagatcca cagaaactta aaacacccgc 720
tgatatcatg ccgctctacc tctggctgat gggcgatgac agccgccgta aaaccggcat 780
gacctttgac gcccaaccgg gccgtaaacc aggaatttcc caatgagtga tgaacgctac 840
caacagcgtc agcagcgagt gaaagaaaaa gtagatgctc gtgtggccca ggcccaggat 900
gaacgcggta ttatcatcgt ctttaccggc aatggaaaag gcaaaaccac cgcggcattt 960
ggtacggcaa cacgcgcagt tggtcacgga aaaaaagtag gcgtcgtgca gtttattaaa 1020
ggcacctggc ctaatggcga acgcaatctg ctggagccac atggcgttga gtttcaggtg 1080
atggcaacgg gctttacctg ggatacacaa aaccgcgagt ctgataccgc cgcctgccgc 1140
gaagtctggc aacatgcaaa gcggatgctt gctgattcct cactggatat ggttttgctt 1200
gatgaactga cgtatatggt ggcgtatgac tatttgccac tggaagaagt ggtgcaggcg 1260
ttaaatgaac gtccacatca acagacggtg attatcacgg gtcgtggttg tcatcgggat 1320
attcttgaac tggcagatac ggtaagtgaa ttacgccccg tcaaacatgc gtttgatgcc 1380
ggtgtaaaag cgcagatagg gatcgattat taatctgtaa aaatgccgga tatgacgtcc 1440
ggcatacttt gattaacctc gagtat 1466

<210> 7
<211> 1247
<212> DNA
<213> Plasmid

<400> 7
gacggtatcg ataagctgga tccataactt cgtataatgt atgctatacg aagttatcta 60
gagtccgaat aaatacctgt gacggaagat cacttcgcag aataaataaa tcctggtgtc 120
cctgttgata ccgggaagcc ctgggccaac ttttggcgaa aatgagacgt tgatcggcac 180
gtaagaggtt ccaactttca ccataatgaa ataagatcac taccgggcgt atttttgag 240
ttatcgagat tttcaggagc taaggaagct aaaatggaga aaaaaatcac tggatatacc 300
accgttgata tatcccaatg gcatcgtaaa gaacattttg aggcatttca gtcagttgct 360
caatgtacct ataaccagac cgttcagctg gatattacgg ccttttaaa gaccgtaaag 420

aaaaataagc acaagtttta tccggccttt attcacattc ttgcccgcct gatgaatgct 480
catccggaat tccgtatggc aatgaaagac ggtgagctgg tgatatggga tagtgttcac 540
ccttgttaca ccgttttcca tgagcaaact gaaacgtttt catcgctctg gagtgaatac 600
cacgacgatt tccggcagtt tctacacata tattcgcaag atgtggcgtg ttacggtgaa 660
aacctggcct atttccctaa agggtttatt gagaatatgt ttttcgtctc agccaatccc 720
tgggtgagtt tcaccagttt tgatttaaac gtggccaata tggacaactt cttcgccccc 780
gttttcaccc atgggcaaat attatacgca aggcgacaag gtgctgatgc cgctggcgat 840
tcaggttcat catgccgtct gtgatggctt ccatgtcggc agaatgctta atgaattaca 900
acagtactgc gatgagtggc agggcggggc gtaatttttt taaggcagtt attggtgccc 960
ttaaacgcct ggtgctacgc ctgaataagt gataataagc ggatgaatgg cagaaattcg 1020
gacctgcagg catgcaactc tagataactt cgtataatgt atgctatacg aagttatgcg 1080
gccgccatat gcatcctagg cctattaata ttccggagta tacgtagccg gctaacgttc 1140
tagcatgcga aatttaaagc gctgatatcg atcgcgcgca gatctgtcat gatgatcatt 1200
gcaattggat ccatatatag ggcccggggt tataattacc tcaggtc 1247

<210> 8
<211> 252
<212> PRT
<213> Escherichia coli

<400> 8

```
Met His Tyr Gln Pro Lys Gln Asp Leu Leu Asn Asp Arg Ile Ile Leu
  1               5                  10                  15

Val Thr Gly Ala Ser Asp Gly Ile Gly Arg Glu Ala Ala Met Thr Tyr
             20                  25                  30

Ala Arg Tyr Gly Ala Thr Val Ile Leu Leu Gly Arg Asn Glu Glu Lys
         35                  40                  45

Leu Arg Gln Val Ala Ser His Ile Asn Glu Glu Thr Gly Arg Gln Pro
     50                  55                  60

Gln Trp Phe Ile Leu Asp Leu Leu Thr Cys Thr Ser Glu Asn Cys Gln
 65                  70                  75                  80

Gln Leu Ala Gln Arg Ile Ala Val Asn Tyr Pro Arg Leu Asp Gly Val
                 85                  90                  95

Leu His Asn Ala Gly Leu Leu Gly Asp Val Cys Pro Met Ser Glu Gln
            100                 105                 110

Asn Pro Gln Val Trp Gln Asp Val Met Gln Val Asn Val Asn Ala Thr
        115                 120                 125

Phe Met Leu Thr Gln Ala Leu Leu Pro Leu Leu Leu Lys Ser Asp Ala
    130                 135                 140

Gly Ser Leu Val Phe Thr Ser Ser Ser Val Gly Arg Gln Gly Arg Ala
145                 150                 155                 160

Asn Trp Gly Ala Tyr Ala Ala Ser Lys Phe Ala Thr Glu Gly Met Met
                165                 170                 175

Gln Val Leu Ala Asp Glu Tyr Gln Gln Arg Leu Arg Val Asn Cys Ile
            180                 185                 190

Asn Pro Gly Gly Thr Arg Thr Ala Met Arg Ala Ser Ala Phe Pro Thr
        195                 200                 205

Glu Asp Pro Gln Lys Leu Lys Thr Pro Ala Asp Ile Met Pro Leu Tyr
    210                 215                 220
```

Leu Trp Leu Met Gly Asp Asp Ser Arg Arg Lys Thr Gly Met Thr Phe
225 230 235 240

Asp Ala Gln Pro Gly Arg Lys Pro Gly Ile Ser Gln
245 250

<210> 9
<211> 22
<212> DNA
<213> primer

<400> 9
cagacaacaa ttccacgagt cc 22

<210> 10
<211> 22
<212> DNA
<213> primer

<400> 10
tgaagttgaa cgtgaatatg cc 22

<210> 11
<211> 28
<212> DNA
<213> primer

<400> 11
tcaaacccgg tggtttctcg cgaccggg 28

<210> 12
<211> 28
<212> DNA
<213> primer

<400> 12
ctcagccgga tatcgacggc gcgctggt 28

<210> 13
<211> 60
<212> DNA
<213> primer

<400> 13
accagcgcgc cgtcgatatc cggctgagta ctcaacacct gccagctctt tacgcaggtt 60

<210> 14
<211> 28
<212> DNA
<213> Primer

<400> 14
cagcatgcct gcgaaccaca ggcctatc 28

<210> 15
<211> 28
<212> DNA
<213> Primer

<400> 15
atgaacaagt ggggcgtagg gttaacat 28

<210> 16
<211> 28

<212> DNA
<213> Primer

<400> 16
ttaattactt gatttattgt cggcttta 28

<210> 17
<211> 2095
<212> DNA
<213> Escherichia coli

<400> 17
cagacaacaa ttccacgagt ccccaccttg gttaaacaac gctttggctc acagtttaca 60
acaggcatgt taagccgtct cgcagagatg ttattgcagc gaaagtatga caaagtatcc 120
attttgcggt gttatcgccc caggctgatt gaaatgccct tctgtttcag gcataaagcc 180
ccaaagtcat aaagtacact ggcagcgcga cagacaacgt cgttgaggag tagccatgca 240
ttaccagcca aaacaagatt tactcaatga tcgcattatc ctggtgacgg gagccagcga 300
tggtattggt cgtgaagccg cgatgacgta tgcacgctat ggtgcgacag tgattctgtt 360
gggccgtaat gaagaaaaat tacgtcaggt agccagccac ataaacgaag aaactgggcg 420
tcagccacag tggtttattc tcgatttgct gacctgcacg tccgaaaatt gccaacaact 480
ggcacagcgc attgccgtta attatccgcg tctggatggt gttttgcata atgccggatt 540
gctcggcgat gtttgcccaa tgagcgaaca aaatccgcag gtctggcagg acgtcatgca 600
ggtcaacgtt aatgccacct ttatgctcac ccaggcactg cttcctttat tactcaaatc 660
ggacgccggt tcactggtct ttacttcatc aagcgttgga cgtcagggac gagccaactg 720
gggtgcatat gcagcgtcga aatttgccac cgaagggatg atgcaggtac tggccgatga 780
atatcagcag cgcctgcgtg tcaactgcat taacccaggc ggtacgcgca ccgcaatgcg 840
tgccagcgcc ttcccgaccg aagatccaca gaaacttaaa acacccgctg atatcatgcc 900
gctctacctc tggctgatgg gcgatgacag ccgccgtaaa accggcatga cctttgacgc 960
ccaaccgggc cgtaaaccag gaatttccca atgagtgatg aacgctacca acagcgtcag 1020
cagcgagtga aagaaaaagt agatgctcgt gtggcccagg cccaggatga acgcggtatt 1080
atcatcgtct ttaccggcaa tggaaaaggc aaaaccaccg cggcatttgg tacggcaaca 1140
cgcgcagttg gtcacggaaa aaaagtaggc gtcgtgcagt ttattaaagg cacctggcct 1200
aatggcgaac gcaatctgct ggagccacat ggcgttgagt ttcaggtgat ggcaacgggc 1260
tttacctggg atacacaaaa ccgcgagtct gataccgccg cctgccgcga agtctggcaa 1320
catgcaaagc ggatgcttgc tgattcctca ctggatatgg ttttgcttga tgaactgacg 1380
tatatggtgg cgtatgacta tttgccactg gaagaagtgg tgcaggcgtt aaatgaacgt 1440
ccacatcaac agacggtgat tatcacgggt cgtggttgtc atcgggatat tcttgaactg 1500
gcagatacgg taagtgaatt acgccccgtc aaacatgcgt ttgatgccgg tgtaaaagcg 1560
cagataggga tcgattatta atctgtaaaa atgccggata tgacgtccgg catactttga 1620
ttaaccgtta ttattacgtc cgccagaacg acgaccgccg ctcacctgac tgtgacgttt 1680
caccgcacga cgaatctgat tcgccttcat gcgacgacgg tctttttcta cagcgacttt 1740
agagctggtt tctggcggta gctccaccag ttcgcgcaga tagttagtct gggcgaggtc 1800
cagctccgtc cagccaccgc gtggcagacc ttttggcagt gggatatcac cgtaacgaac 1860
acggatcagg cggctcacct gcacaccaac cgcttcccac agacgacgaa cttcacggtt 1920
acggccttcg gtcagagtca cgttgtacca ctggttgatc ccttcgccgc cgctgaactt 1980
gatggtttta aaagctgccg gaccatcttc caactgcacg ccacgactca aatcacgcag 2040
tttcgcgtcg tcaacctgac caaatacacg cacggcatat tcacgttcaa cttca 2095

<210> 18
<211> 969
<212> DNA
<213> Escherichia coli

<400> 18
ggtggcggcc gctctagccc ataaagcccc aaagtcataa agtacactgg cagcgcgaca 60
gacaacgtcg ttgaggagta gccatgcatt accagccaaa acaagattta ctcaatgatc 120
gcattatcct ggtgacggga gccagcgatg gtattggtcg tgaagccgcg atgacgtatg 180
cacgctatgg tgcgacagtg attctgttgg gccgtaatga agaaaaatta cgtcaggtag 240
ccagccacat aaacgaagaa actgggcgtc agccacagtg gtttattctc gatttgctga 300
cctgcacgtc cgaaaattgc caacaactgg cacagcgcat tgccgttaat tatccgcgtc 360
tggatggtgt tttgcataat gccggattgc tcggcgatgt ttgcccaatg agcgaacaaa 420
atccgcaggt ctggcaggac gtcatgcagg tcaacgttaa tgccaccttt atgctcaccc 480
aggcactgct tcctttatta ctcaaatcgg acgccggttc actggtcttt acttcatcaa 540
gcgttggacg tcagggacga gccaactggg gtgcatatgc agcgtcgaaa tttgccaccg 600

aagggatgat gcaggtactg gccgatgaat atcagcagcg cctgcgtgtc aactgcatta 660
acccaggcgg tacgcgcacc gcaatgcgtg ccagcgcctt cccgaccgaa gatccacaga 720
aacttaaaac acccgctgat atcatgccgc tctacctctg gctgatgggc gatgacagcc 780
gccgtaaaac cggcatgacc tttgacgccc aaccgggccg taaaccagga atttcccaat 840
gagtgatgaa cgctaccaac agcgtcagca gcgagtgaaa gaaaaagtag atgctcgtgt 900
ggcccaggcc caggatgaac gcggtattat catcgtcttt accggcaatg gaaaaggcaa 960
aaccaccgc 969

<210> 19
<211> 1282
<212> DNA
<213> Escherichia coli

<400> 19
ggccagttct tctttcatct ccttatactg ttcgctgaga ttgttgaccc gtaactcgcc 60
acgctggcgt ttattacgtt gagcaacatt gacaataatg gcggcaatcg ccgcaatcgc 120
tagcacaacg gtaacgattt tcgccaaaaa caaaccatat tcagacaaca attccacgag 180
tccccacctt ggttaaacaa cgctttggct cacagtttac aacaggcatg ttaagccgtc 240
tcgcagagat gttattgcag cgaaagtatg acaaagtatc cattttgcgg tgttatcgcc 300
ccaggctgat tgaaatgccc ttctgtttca ggcataaagc cccaaagtca taaagtacac 360
tggcagcgcg acagacaacg tcgttgagga gtagccatgc attaccagcc aaaacaagat 420
ttactcaatg atcgcattat cctggtgacg ggagccagcg atggtattgg tcgtgaagcc 480
gcgatgacgt atgcacgcta tggtgcgaca gtgattctgt tgggccgtaa tgaagaaaaa 540
ttacgtcagg tagccagcca cataaacgaa gaaactgggc gtcagccaca gtggtttatt 600
ctcgatttgc tgacctgcac gtccgaaaat tgccaacaac tggcacagcg cattgccgtt 660
aattatccgc gtctggatgg tgttttgcat aatgccggat tgctcggcga tgtttgccca 720
atgagcgaac aaaatccgca ggtctggcag gacgtcatgc aggtcaacgt taatgccacc 780
tttatgctca cccaggcact gcttccttta ttactcaaat cggacgccgg ttcactggtc 840
tttacttcat caagcgttgg acgtcaggga cgagccaact ggggtgcata tgcagcgtcg 900
aaatttgcca ccgaagggat gatgcaggta ctggccgatg aatatcagca gcgcctgcgt 960
gtcaactgca ttaacccagg cggtacgcgc accgcaatgc gtgccagcgc cttcccgacc 1020
gaagatccac agaaacttaa aacacccgct gatatcatgc cgctctacct ctggctgatg 1080
ggcgatgaca gccgccgtaa aaccggcatg acctttgacg cccaaccggg ccgtaaacca 1140
ggaatttccc aatgagtgat gaacgctacc aacagcgtca gcagcgagtg aaagaaaaag 1200
tagatgctcg tgtggcccag gcccaggatg aacgcggtat tatcatcgtc tttaccggca 1260
atggaaaagg caaaaccacc gc 1282

INTERNATIONAL SEARCH REPORT

Internatic Application No
PCT/US 00/17154

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12N15/90 C12N15/67

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 23238 A (AMGEN INC) 14 May 1999 (1999-05-14) see whole document, particularly page 5, lines 21-22, page 7, lines 18-19, page 8, lines 11-13, and page 20, line 13. ---	1-8
X	WO 98 42727 A (STANFORD RES INST INT) 1 October 1998 (1998-10-01) See whole document, particularly the abstract, page 19, lines 21-22, page 25, lines 13-16, page 44, lines 19-21, and page 50, lines 5-7. --- -/--	1-8

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 October 2000	24/10/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Smalt, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internatio Application No
PCT/US 00/17154

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	MURPHY, K.C.: "Use of bacteriophage lambda recombination funtions to promote gene replacement in Escherichia coli." JOURNAL OF BACTERIOLOGY, vol. 180, no. 8, April 1998 (1998-04), pages 2063-71, XP002149589	1-6,8
Y	the whole document	7
Y	WO 90 12098 A (CETUS CORP) 18 October 1990 (1990-10-18) page 5, line 28 - line 30; example XVI page 17, line 3 - line 5 page 69, line 5 - line 6 page 74, line 17 - line 29	7
X	SHIROZA T ET AL: "Application of the resident plasmid integration technique to construct a strain of Streptococcus godronii able to express the Bacillus circulans cycloisomaltooligosaccharide glucanotransferase gene, and secrete its active gene product" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES,GB,ELSEVIER SCIENCE PUBLISHERS, BARKING, vol. 207, no. 2, 30 January 1998 (1998-01-30), pages 119-126, XP004108909 ISSN: 0378-1119 the whole document	1-6
X	US 5 530 191 A (MALIGA PAL) 25 June 1996 (1996-06-25) column 24, line 13 - line 16	7
X	WHITE A P ET AL: "High efficiency gene replacement in Salmonella enteritidis: chimeric fimbrins containing a T-cell epitope from Leishmania major" VACCINE,GB,BUTTERWORTH SCIENTIFIC. GUILDFORD, vol. 17, no. 17, 23 April 1999 (1999-04-23), pages 2150-2161, XP004165000 ISSN: 0264-410X the whole document	1-6,8
X	EL KAROUI MERIEM ET AL: "Gene replacement with linear DNA in electroporated wild-type Escherichia coli." NUCLEIC ACIDS RESEARCH, vol. 27, no. 5, 1 March 1999 (1999-03-01), pages 1296-1299, XP002149590 ISSN: 0305-1048 the whole document	1-6

-/--

INTERNATIONAL SEARCH REPORT

Internatio Application No
PCT/US 00/17154

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 837 528 A (PERO JANICE G ET AL) 17 November 1998 (1998-11-17) column 12, line 47 - line 51; example 8	7
P,X	WO 99 66035 A (ICONIX PHARMACEUTICALS INC) 23 December 1999 (1999-12-23) the whole document	1-6,8

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internatio· \pplication No
PCT/US 00/17154

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9923238 A	14-05-1999	US 6090629 A	18-07-2000
		AU 7606598 A	24-05-1999
		EP 1030930 A	30-08-2000
WO 9842727 A	01-10-1998	US 5948653 A	07-09-1999
		AU 6562098 A	20-10-1998
		EP 0977771 A	09-02-2000
		US 6074853 A	13-06-2000
WO 9012098 A	18-10-1990	AT 161049 T	15-12-1997
		AU 5437390 A	05-11-1990
		CA 2014264 A	12-10-1990
		DE 69031803 D	22-01-1998
		DE 69031803 T	09-07-1998
		DK 471687 T	24-08-1998
		EP 0471687 A	26-02-1992
		ES 2111536 T	16-03-1998
		JP 7073500 B	09-08-1995
		NZ 233312 A	26-03-1992
		US 5268274 A	07-12-1993
US 5530191 A	25-06-1996	AU 687544 B	26-02-1998
		AU 2198195 A	09-10-1995
		BR 9507180 A	14-10-1997
		CA 2186164 A	28-09-1995
		CN 1148869 A	30-04-1997
		EP 0805852 A	12-11-1997
		JP 9510615 T	28-10-1997
		NZ 283622 A	19-12-1997
		PL 316435 A	06-01-1997
		RU 2139347 C	10-10-1999
		WO 9525787 A	28-09-1995
US 5837528 A	17-11-1998	US 5925538 A	20-07-1999
		AT 195971 T	15-09-2000
		CN 1049185 A	13-02-1991
		DE 69033616 D	05-10-2000
		EP 0405370 A	02-01-1991
		EP 1001026 A	17-05-2000
		JP 3117489 A	20-05-1991
		JP 10066562 A	10-03-1998
WO 9966035 A	23-12-1999	AU 4694399 A	05-01-2000